

RESTRICTION MAP OF THE RSV F GENE

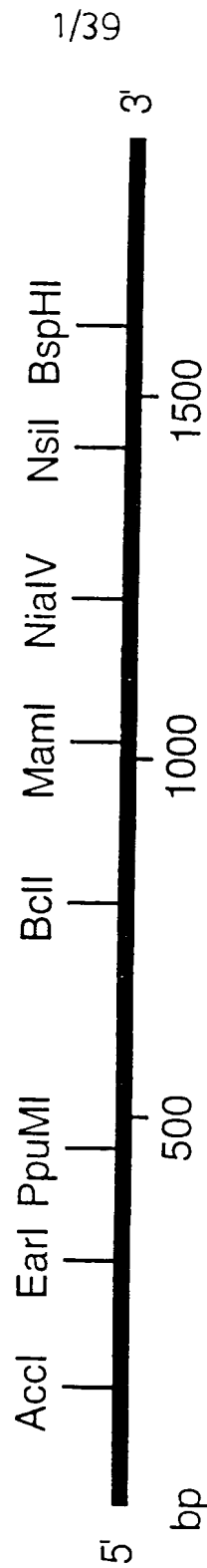


FIG.1

19/9/4621

FIG.2A.

NUCLEOTIDE SEQUENCE OF THE RSV F GENE.

5' MET GLU LEU PRO ILE LEU LYS ALA ASN ALA ILE THR THR ILE LEU ALA VAL THR PHE  
 ATGGAGTTGCCAATCCTCAAGCAAAATGCAATACCACAATCCTCGCTGCAGTCACATTT  
 TACCTCAACGGTTAGGAGTTTCGTTTACGTTAATGGTGTAGGAGCGACGTCAGTGTA  
 10 20 30 40 50 60  
 CYS PHE ALA SER SER GLN ASN ILE THR GLU GLU PHE TYR GLN SER THR CYS SER ALA VAL  
 TGCTTTGCTTCTAGTCAAAACATCACTGAAAGAAATTTTATCAATCAACATGCAGTGCAGTT  
 ACGAAACGAAGATCAGTTTGTAGTGACTTCTTAAATAAGTTAGTTGTACGTCACGTCAA  
 70 80 90 100 110 120  
 SER LYS GLY TYR LEU SER ALA LEU ARG THR GLY TRP TYR THR SER VAL ILE THR ILE GLU  
 AGCAAAGGCTATCTTAGTGCTCTAAGAACTGGTTGGTATACTAGTGTATACTATAGAA  
 TCGTTTCCGATAGAAATCAGGAGATTCTTGACCAACCATATGATCACAATATTGATATCTT  
 130 140 150 160 170 180  
 LEU SER ASN ILE LYS GLU ASN LYS CYS ASN GLY THR ASP ALA LYS VAL LYS LEU MET LYS  
 TTAAGTAATATCAAGGAAATTAAGTGTAAATGGAACAGATGCTAAGGTAATAATTGATGAAA  
 AATTCATTATAGTTCCTTTTATTACATTTACCTTGCTCTACGATTCCATTTTAACTACTTT  
 190 200 210 220 230 240  
 GLN GLU LEU ASP LYS TYR LYS ASN ALA VAL THR GLU LEU GLN LEU MET GLN SER THR  
 CAAGAAATTAGATAAATAAATAATGCTGTAAACAGAATTGCAGTTGCTCATGCAAAGCACA  
 GTTCTTAATCTATTATATTTTACGACATTGTCTTAACGTCAACGAGTACGTTTCGTGT  
 250 260 270 280 290 300  
 PRO ALA ALA ASN ASN ARG ALA ARG ARG GLU LEU PRO ARG PHE MET ASN TYR THR LEU ASN  
 CCAGCAGCAAACAATCGAGCCAGAGAGAACTACCAAGGTTTATGAATTATACACTCAAC  
 GGTCGTCGTTTGTAGCTCGGCTCTTCTTGATGGTTCCAAATACTTAATATGTGAGTTG  
 310 320 330 340 350 360

3/39

FIG.2B.

F2-F1CLEAVAGE SITE

ASN THR LYS LYS THR ASN VAL THR LEU SER LYS LYS ARG LYS ARG ARG PHE LEU GLY PHE  
 AATACCAAAAACCAATGTAACATTAAGCAAGAAAGGAA8AAGATTCTTGTTT  
 TTATGGTTTTGTGTTACATTTGTAATTCGTTCTTTTCTTTCTTCTAAAGAACCAAAA  
 370 380 390 400 410 420

LEU LEU GLY VAL GLY SER ALA ILE ALA SER GLY ILE ALA VAL SER LYS VAL LEU HIS LEU  
 TTGTTAGGTGTTGGATCTGCAATCGCCAGTGGCATTTGCTGTATCTAAGGTCCTGCACCTTA  
 AACAAATCCACAACCTAGACGTTAGCGGTACCGTACCGACATAGATTCCAGGACGTGAAT  
 430 440 450 460 470 480

GLU GLY GLU VAL ASN LYS ILE LYS SER ALA LEU LEU SER THR ASN LYS ALA VAL VAL SER  
 GAAGGAGAGTGAACAAGATCAAAAGTGCTCTACTATCCACAACAAGCCGTAGTCAGC  
 CTTCCTCTTCACTTGTTCTAGTTTTCACCGAGATGATAGGTGTTTGTTCGGCATCAGTCG  
 490 500 510 520 530 540

LEU SER ASN GLY VAL SER VAL LEU THR SER LYS VAL LEU ASP LEU LYS ASN TYR ILE ASP  
 TTATCAAAATGGAGTTAGTGCTTTAACCAAGCAAGTGTAGACCTCAAAAACATATATAGAT  
 AATAGTTTACCTCAATCACAGAAATTGGTCGTTTCACAATCTGGAGTTTGTGATATATCTA  
 550 560 570 580 590 600

LYS GLN LEU LEU PRO ILE VAL ASN LYS GLN SER CYS ARG ILE SER ASN ILE GLU THR VAL  
 AAACAATTGTACCTATTGTGAATAAGCAAGCTGCAGAAATATCAAAATATAGAAACTGTG  
 TTTGTTAAACAATGGATAACACTTATTTCGTTTCGACGCTTATAGTTTATATCTTTTGACAC  
 610 620 630 640 650 660

ILE GLU PHE GLN HIS LYS ASN ASN ARG LEU LEU GLU ILE THR ARG GLU PHE SER VAL ASN  
 ATAGAGTTCCACAACAAGAACACAGACTACTAGAGATTACCAGGGAATTAGTGTAAAT  
 TATCTCAAGGTGTTTCTTTGTTGCTGATGATCTCTAATGGTCCCTTAAATCACAAATTA  
 670 680 690 700 710 720

ALA GLY VAL THR THR PRO VAL SER THR TYR MET LEU THR ASN SER GLU LEU LEU SER LEU  
 GCAGGTGTAACACTACACCTGTAGCACTTACATGTTAACTAATAGTGAATTTATGTCATTA  
 CGTCCACATTGATGTGGACATTCTGTAATGTACAATTGATTATCACTTAATAACAGTAAT  
 730 740 750 760 770 780

09/9/4621

4/39

## FIG.2C.

```

ILE ASN ASP MET PRO ILE THR ASN ASP GLN LYS LYS LEU MET SER ASN ASN VAL GLN ILE
ATCAATGATATGCCTATAACAATGATCAGAAAAGTTAATGTCCAACAATGTTCAAATA
TAGTTACTATACGGATATTGTTTACTAGTCTTTTCAATTACAGGTGTTACAAAGTTTAT
790      800      810      820      830      840

VAL ARG GLN GLN SER TYR SER ILE MET SER ILE ILE LYS GLU VAL LEU ALA TYR VAL
GTTAGACAGCAAGTTACTCTATCATGTCCATAATAAAAGAGGAAGTCTTAGCATATGTA
CAATCTGTCGTTTCAATGAGATAGTACAGGTATTATTCTCTCCTTCAGAATCGTATACAT
850      860      870      880      890      900

VAL GLN LEU PRO LEU TYR GLY VAL ILE ASP THR PRO CYS TRP LYS LEU HIS THR SER PRO
GTACAATTACCACCTATATGGTGTGATAGATACACCTTGTGGAAATTACACACAT'CCCCCT
CATGTTAATGGTGATATACCACACTATCTATGTGGAACAACCTTTAATGTGTAGGGGA
910      920      930      940      950      960

LEU CYS THR THR ASN THR LYS GLU GLY SER ASN ILE CYS LEU THR ARG THR ASP ARG GLY
CTATGTACAACCAACACAAAGAGGGTCAACATCTGTTTAACAAGAACTGACAGAGGA
GATACATGTTGGTGTGTTTCTTCCAGTTTGTAGACAAAATTGTTCTTGACTGTCTCCT
970      980      990      1000      1010      1020

TRP TYR CYS ASP ASN ALA GLY SER VAL SER PHE PHE PRO GLN ALA GLU THR CYS LYS VAL
TGGTACTGTGACAAATGCAGGATCAGTATCTTTCTTCCACAAGCTGAAACATGTAAAGTT
ACCATGACACTGTTACGTCCTAGTCATAGAAAGAGGGTGTTCGACTTTGTACATTTTCAA
1030      1040      1050      1060      1070      1080

GLN SER ASN ARG VAL PHE CYS ASP THR MET ASN SER LEU THR LEU PRO SER GLU VAL ASN
CAATCGAATCGAGTATTTGTGACACAATGAACAGTTTAAACATTACCAAGTGAAGTAAAT
GTTAGCTTAGCTCATAAAACACTGTGTACTTGTCAAATTTGTAATGGTTCACCTTCATTTA
1090      1100      1110      1120      1130      1140

LEU CYS ASN VAL ASP ILE PHE ASN PRO LYS TYR ASP CYS LYS ILE MET THR SER LYS THR
CTCTGCAATGTTGACATATTCAAATCCCAAATATGATTGTAAAATTATGACTTCAAAAACA
GAGACGTTACAACTGTATAAGTTAGGGTTTATATACTAACATTTTAATACTGAAGTTTGTGT
1150      1160      1170      1180      1190      1200

```

19/9/4621

ASP VAL SER SER SER VAL ILE THR SER LEU GLY ALA ILE VAL SER CYS TYR GLY LYS THR  
GATGTAAGCAGCTCCGTTATTCACATCTCTAGGAGCCATTGTGTCATGCTATGGCAAAACT  
CTACATTCGTCGAGGCAATAGTGTAGAGATCCCTCGGTAAACACAGTACGATACCGTTTGA  
1210 1220 1230 1240 1250 1260

LYS CYS THR ALA SER ASN LYS ASN ARG GLY ILE ILE LYS THR PHE SER ASN GLY CYS ASP  
AAATGTAACAGCATCCCAATAAATAATCGTGGAATCATAAAGACATTTCTTAACGGGTGTGAT  
TTTACATGTCGTAGGTTATTTTATAGCACCTTAGTATTTCTGTAAAGATTGCCCCACACTA  
1270 1280 1290 1300 1310 1320

TYR VAL SER ASN LYS GLY VAL ASP THR VAL SER VAL GLY ASN THR LEU TYR TYR VAL ASN  
TATGTAATCAAAATAAAGGGGTGGACACTGTGTCGTAGGTAAACACATTTATATATGTAAAT  
ATACATAGTTTATTTCCCCACCTGTGACACAGACATCCCATTTGTGTAAATATATACATTTA  
1330 1340 1350 1360 1370 1380

LYS GLN GLU GLY LYS SER LEU TYR VAL LYS GLY GLU PRO ILE ILE ASN PHE TYR ASP PRO 5/39  
AAGCAAGAGGCAAAAGTCTCTATGTAAAGGTGAACCAATAATAAATTTCTATGACCCCA  
TTTCGTTCTTCCGTTTTCAGAGATACATTTTCCACTTGGTTATTTAAAGATACTGGGT  
1390 1400 1410 1420 1430 1440

LEU VAL PHE PRO SER ASP GLU PHE ASP ALA SER ILE SER GLN VAL ASN GLU LYS ILE ASN  
TTAGTATTTCCCTCTCTGATGAAATTTTGATGCAATCAATATCTCAAGTCAATGAGAAAGATTAAAC  
AATCATAGGGGAGACTACTTAAACTACGTAGTTATAGAGTTTCAGTTACTCTTCTAATTG  
1450 1460 1470 1480 1490 1500

GLN SER LEU ALA PHE ILE ARG LYS SER ASP GLU LEU LEU HIS ASN VAL ASN ALA GLY LYS  
CAGAGTTTAGCATTATTCGTAAATCCGATGAATTTATTTACATAAATGTAATAATGCTGGTAA  
GTCTCAATTCGTAAATAAGCATTTAGGCTACTTAATAATGTAATTCATTTACGACCATTT  
1510 1520 1530 1540 1550 1560

SER THR THR ASN ILE MET ILE THR THR ILE ILE GLU ILE ILE VAL ILE LEU LEU SER  
TCAACCAACAATATCATGATGATAACTATAATTTAGAGATTTATAGTAATAATTTGTTATCA  
AGTTGGTGTATATAGTACTATTTGATGATATTAATCTCTAATATCATTAATAACAATAGT  
1570 1580 1590 1600 1610 1620

FIG.2D.

09/9/14621

LEU ILE ALA VAL GLY LEU LEU LEU TYR CYS LYS ALA ARG SER THR PRO VAL THR LEU SER  
 TTAATTGGCTGTGGACTGGCTCCTATACCTGTAAGGCCAGAAAGCACACCAAGTCAACACTAAGC  
 AATTAAACGACAACTGACGAGGATATGACATTCCTGGTCTTCGTGTTCAATGTTGATTCG  
 1630 1640 1650 1660 1670 1680  
 LYS, ASP GLN LEU SER GLY ILE ASN ASN ILE ALA PHE SER ASN  
 AAGGATCAACTGAGTGGTATAAATAATATTTGCAATTTAGTAACCTGAATAAATAAGCACCT  
 TTCCTAGTTGACTCACCATATTTATTAATAACGTAAATTCATTGACTTATTTTATCCTGTTGGA  
 1690 1700 1710 1720 1730 1740  
 AATCATGTTCTTACAAATGGTTTACTATCTGCTCATAGACAACCCATCTATCATTTGGATT  
 TTAGTACAAGAATGTTACCAAAATGATAGACGAGTATCTGTTGGGTAGATAGTAACCTAAA  
 1750 1760 1770 1780 1790 1800  
 TCTTAAATAATCTGAACCTTCAATCGAAACTCTTATCTATAAACCATCTCACTTACACTATTTA  
 AGAATTTTAGACTTGAAGTAGCTTTGAGAAATAGATATTTGTTAGAGTGAATGTGATAAAT  
 1810 1820 1830 1840 1850 1860  
 AGTAGATTCCCTAGTTTATAGTTATAT 3'  
 TCATCTAAGGATCAAAATATCAATATA  
 1870 1880

NUCLEOTIDE SEQUENCE OF THE RSV F GENE. THE CDNA SEQUENCE IS SHOWN IN THE PLUS (mRNA)  
 STRAND SENSE IN THE 5' TO 3' DIRECTION. THE SIGNAL PEPTIDE (SP) AND THE TRANSMEMBRANE (TM)  
 ANCHOR DOMAIN ARE UNDERLINED. THE PREDICTED F2-F1 CLEAVAGE SITE IS INDICATED BY THE ARROW  
 (↓).

FIG. 2e

89/9/4621

7/39

FIG. 3A.

NUCLEOTIDE SEQUENCE OF THE RSV F GENE.

5' ↓  
 MET GLU LEU PRO ILE LEU LYS ALA ASN ALA ILE THR THR ILE LEU ALA VAL THR PHE  
 ATGGAGTTGCCAATCCTCAAGCAAAATGCAATTACCAATCCTCGCTGCAGTCACATTT  
 TACCTCAACGGTTAGGAGTTTCGTTTACGTTAATGGTGTAGGAGCGACGTCAGTGTA  
 10 20 30 40 50 60  
 CYS PHE ALA SER SER GLN ASN ILE THR GLU GLU PHE TYR GLN SER THR CYS SER ALA VAL  
 TGCTTTGCTTCTAGTCAAAACATCACTGAAGAATTTTATCAATCAACATGCAGTGCAGTT  
 ACGAAACGGAAGATCAGTTTGTAGTGACTTCTTAAATAAGTTAGTTGTACGTCACGTCAA  
 70 80 90 100 110 120  
 SER LYS GLY TYR LEU SER ALA LEU ARG THR GLY TRP TYR THR SER VAL ILE THR ILE GLU  
 AGCAAGGCTATCTTAGTGCTCTAAGAACTGGTTGGTATATACTAGTGTATACTATAGAA  
 TCGTTTCCGATAGAAATCACGAGATTCTTGACCAACCATATGATCACAATATTGATATCTT  
 130 140 150 160 170 180  
 LEU SER ASN ILE LYS GLU ASN LYS CYS ASN GLY THR ASP ALA LYS VAL LYS LEU MET LYS  
 TTAAGTAATATCAAGGAAATAAGTGTAAATGGAACAGATGCTAAGGTAAATTTGATGAAA  
 AATTCAATTATAGTTCCTTTTATTCACATTACCTTGCTACGATTCCATTTTAACTACTTT  
 190 200 210 220 230 240  
 GLN GLU LEU ASP LYS TYR LYS ASN ALA VAL THR GLU LEU GLN LEU MET GLN SER THR  
 CAAGAATTAGATAAATAAATAATGCTGTAACAGAATTGCAGTTGCTCATGCAAGCACACA  
 GTTCTTAATCTATTTATATTTTACGACATTGTCTTAACGTCACGAGTACGTTTCGTGT  
 250 260 270 280 290 300  
 PRO ALA ALA ASN ASN ARG ALA ARG ARG GLU LEU PRO ARG PHE MET ASN TYR THR LEU ASN  
 CCAGCAGCAACAATCGAGCCAGAGAGAACTACCAAGGTTTATGAATTATACACTCAAC  
 GGTCGTCGTTTGTAGCTCGGTCCTCTCTTGATGGTTCCAAATACTTAATATGTGAGTTG  
 310 320 330 340 350 360

09/9/4621

8/39

## FIG.3B.

ASN THR LYS LYS THR ASN VAL THR LEU SER LYS LYS ARG LYS ARG ARG F2-F1CLEAVAGE SITE  
AATACCAAAACCAATGTAACATTAAGCAAGAAAAGGAAAAA8AAGATTTCCTGGTTT  
TTATGGTTTTCCTTACATGTAATTCGTTCTTTTCCTTTCTTAAAGAACCAAAA 370 380 390 400 410 420

LEU LEU GLY VAL GLY SER ALA ILE ALA SER GLY ILE ALA VAL SER LYS VAL LEU HIS LEU  
TTGTTAGGTGTTGGATCTGCAATCGCCAGTGGCATTGTGTATCTAAGGTCTGCACTTA  
AACAAATCCACAACCTAGACGTTAGCGGTACCGTAACGACATAGATTCCAGGACGTGAAT 430 440 450 460 470 480

GLU GLY GLU VAL ASN LYS ILE LYS SER ALA LEU LEU SER THR ASN LYS ALA VAL VAL SER  
GAAGGAGAGTGAACAAGATCAAAAGTGCTCTACTATCCACAACAAGCCGTAGTCAGC  
CTTCCTCTTCACTTGTTCTAGTTTTCACGAGATGATAGGTGTTTGTCCGGCATCAGTCG 490 500 510 520 530 540

LEU SER ASN GLY VAL SER VAL LEU THR SER LYS VAL LEU ASP LEU LYS ASN TYR ILE ASP  
TTATCAAAATGGAGTTAGTGTCTTAAACCAGCAAGTGTAGACCTCAAAAACATATATAGAT  
AATAGTTTACCTCAATCACAGAAATTGGTCGTTTCACAATCTGGAGTTTGTGATATATCTA 550 560 570 580 590 600

LYS GLN LEU LEU PRO ILE VAL ASN LYS GLN SER CYS ARG ILE SER ASN ILE GLU THR VAL  
AAACAATGTACCTATTTGTGAATAAGCAAGCTGCAGAAATATCAAAATATAGAAACTGTG  
TTTGTTAACAATGGATAACACTTATTCGTTTCGACGCTTATAGTTTATATATCTTTTGACAC 610 620 630 640 650 660

ILE GLU PHE GLN HIS LYS ASN ASN ARG LEU LEU GLU ILE THR ARG GLU PHE SER VAL ASN  
ATAGAGTTCACAACAAAGAACACACAGACTACTAGAGATPACCAGGGAATTTAGTGTAAAT  
TATCTCAAGGTGTTTCTTGTGTTCTGATGATCTCTAATGTGTCCTTAAATCACAAATTA 670 680 690 700 710 720

ALA GLY VAL THR THR PRO VAL SER THR TYR MET LEU THR ASN SER GLU LEU LEU SER LEU  
GCAGGTGTAACCTACACCTGTAAAGCATTACATGTAACTAATAGTGAATTAATGTCAATTA  
CGTCCACATTGATGTGGACATTCTGTAATGTACAATTGATTATCACTTAATAACAGTAAT 730 740 750 760 770 780



9/39

## FIG.3C.

ILE ASN ASP MET PRO ILE THR ASN ASP GLN LYS LYS LEU MET SER ASN VAL GLN ILE  
ATCAATGATATGCTTAAACAAATGATCAGAAAAGTTAATGTCCAACAATGTTCAAATA  
TAGTTACTATACGGATATGTTTACTAGTCTTTTTCAAATTACAGGTTGTACAAAGTTTAT  
790 800 810 820 830 840

VAL ARG GLN SER TYR SER ILE MET SER ILE ILE LYS GLU GLU VAL LEU ALA TYR VAL  
GTAGACAGCAAGTTACTCTATCATGTCCATAATAAAGAGGAAGTCTTAGCATATGTA  
CAATCTGTCGTTTCAATGAGATAGTACAGGTATTTTCTCCTTCAGAAATCGTATACAT  
850 860 870 880 890 900

VAL GLN LEU PRO LEU TYR GLY VAL ILE ASP THR PRO CYS TRP LYS LEU HIS THR SER PRO  
GTACAAATTACCACATATGTTGTGATAGATACACCTTGTGGAAATTACACATCCCT  
CATGTTAATGGTGATATACCACACTATCTATGTGGAACAACCTTTAATGTGTAGGGGA  
910 920 930 940 950 960

LEU CYS THR THR ASN THR LYS GLY SER ASN ILE CYS LEU THR ARG THR ASP ARG GLY  
CTATGTACAACCAACACAAAGAGGGTCAACATCTGTTTAAACAAGAACTGACAGAGGA  
GATACATGTTGGTGTGTTTCTTCCAGTTTGTAGACAAATGTTCTTGACTGTCTCCT  
970 980 990 1000 1010 1020

TRP TYR CYS ASP ASN ALA GLY SER VAL SER PHE PRO GLN ALA GLU THR CYS LYS VAL  
TGGTACTGTGACAAATGCAGGATCAGTATCTTCTTCCACAAGCTGAAACATGTAAAGTT  
ACCATGACACTGTTACGTCCTAGTCATAGAAAGGGGTTCGACTTTGTACATTTCAA  
1030 1040 1050 1060 1070 1080

GLN SER ASN ARG VAL PHE CYS ASP THR MET ASN SER LEU THR LEU PRO SER GLU VAL ASN  
CAATCGAATCGAGTATTTTGTGACACAATGAACAGTTTAACATTAACCAAGTGAAGTAAAT  
GTTAGCTTAGCTCATAAACACTGTGTACTTGTCAAATGTAATGGTTCACCTTCAATTA  
1090 1100 1110 1120 1130 1140

LEU CYS ASN VAL ASP ILE PHE ASN PRO LYS TYR ASP CYS LYS ILE MET THR SER LYS THR  
CTCTGCAATGTTGACATATTCATCCCAATATGATGTAAATTTATGACTTCAAAAACA  
GAGACGTTACAACCTGTATAAGTTAGGTTTATATACTAACAATTTTAATACTGAAGTTTGT  
1150 1160 1170 1180 1190 1200

09/9/4621

10/39

## FIG.3D.

ASP VAL SER SER VAL ILE THR SER LEU GLY ALA ILE VAL SER CYS TYR GLY LYS THR  
 GATGTAAGCAGCTCCGTTATCACATCTCTAGGAGCCATTGTGTCATGCTATGGCAAAACT  
 CTACATTCGTCGAGGCAATAGTGTAGAGATCCTCGGTAACACAGTACGATACCGTTTGA  
 1210 1220 1230 1240 1250 1260

LYS CYS THR ALA SER ASN LYS ASN ARG GLY ILE ILE LYS THR PHE SER ASN GLY CYS ASP  
 AAATGTACAGCATCCAATAAATAAATCGTGAATCATAAAGACATTTTCTAACGGGTGTGAT  
 TTTACATGTCTAGGTATTTTGTAGCACCTTAGTATTTCTGTAAAGATTGCCACACTA  
 1270 1280 1290 1300 1310 1320

TYR VAL SER ASN LYS GLY VAL ASP THR VAL SER VAL GLY ASN THR LEU TYR TYR VAL ASN  
 TATGTATCAAAATAAAGGGGTGGACACTGTGCTGTAGGTAACACATTATATATTGTAAT  
 ATACATAGTTTATTTCCCCACCTGTGACACAGACATCCATTGTGTAATATAATACATTTA  
 1330 1340 1350 1360 1370 1380

LYS GLN GLU LYS SER LEU TYR VAL LYS GLY GLU PRO ILE ILE ASN PHE TYR ASP PRO  
 AAGCAAGAGGCAAAAGTCTCTATGTAAAGGTGAACCAATAAJAAATTTCTATGACCCA  
 TTCGTTCTCCGTTTTCAGAGATACATTTTCCACTTGGTTATTTAAAGATACCTGGGT  
 1390 1400 1410 1420 1430 1440

LEU VAL PHE PRO SER ASP GLU PHE ASP ALA SER ILE SER GLN VAL ASN GLU LYS ILE ASN  
 TTAGTATTTCCCTCTGATGAATTTGATGCATCAATATCTCAAGTCAATGAGAAGATTAAAC  
 AATCATAAGGGGAGACTACTTAAACTACGTAGTTATAGAGTTCAGTTACTCTTCTAATTG  
 1450 1460 1470 1480 1490 1500

GLN SER LEU ALA PHE ILE ARG LYS SER ASP GLU LEU LEU HIS ASN VAL ASN ALA GLY LYS  
 CAGAGTTTAGCATTTATTTCGTAAATCCGATGAATTATACATAATGTAATGCTGTGTA  
 GTCTCAAATCGTAAATAAGCATTTAGGCTACTTAATAATGTAATACATTACGACCATTT  
 1510 1520 1530 1540 1550 1560

SER THR THR ASN ILE MET Thr Stop Stop Bam HI  
 TCAACCACAAATATCATGACTTGATAATGAGGATCC  
 AGTTGGTGTATTATAGTACTGAACCTATTACTCCTAGG  
 1570

09/9/4621

11/39

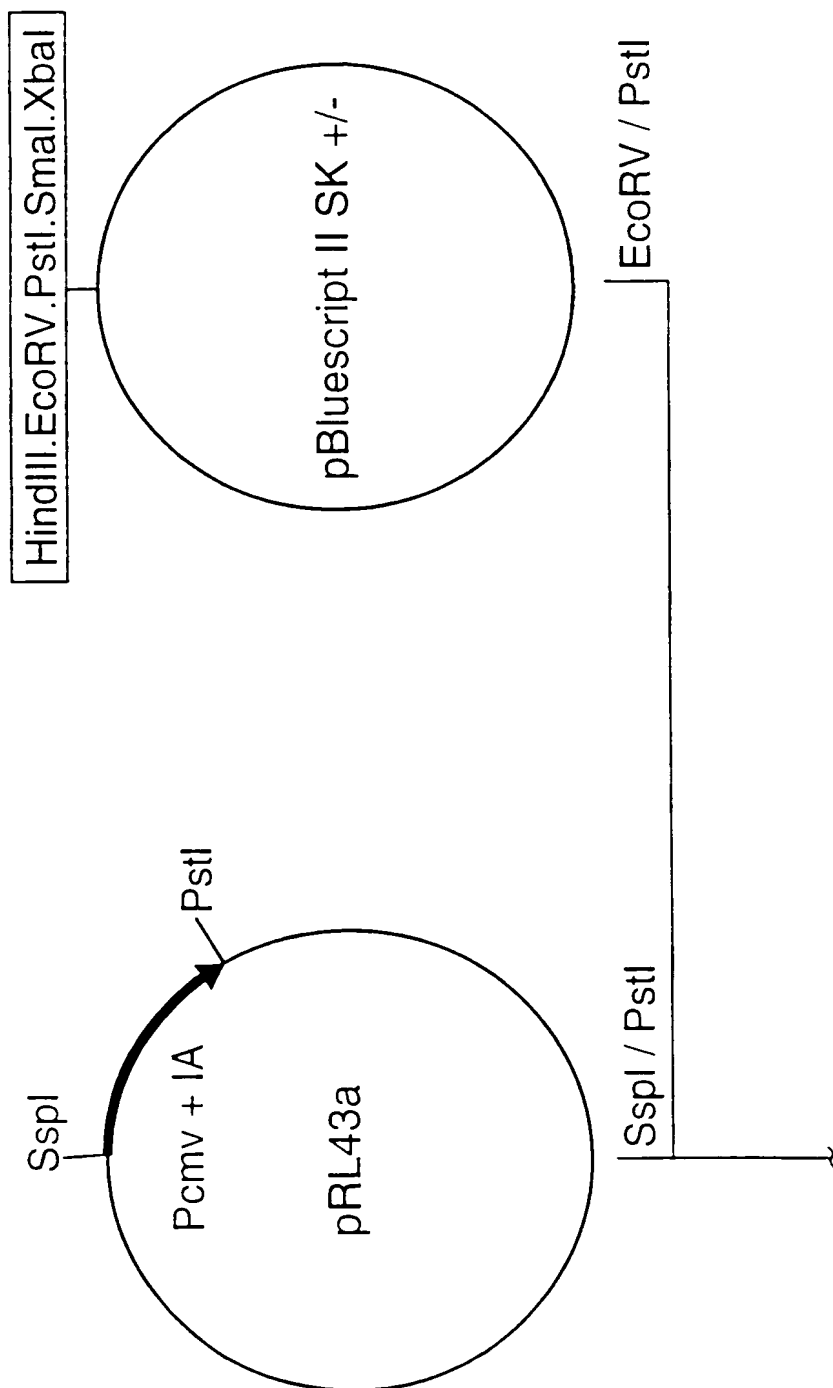


FIG.4A

89/9/4621

12/39

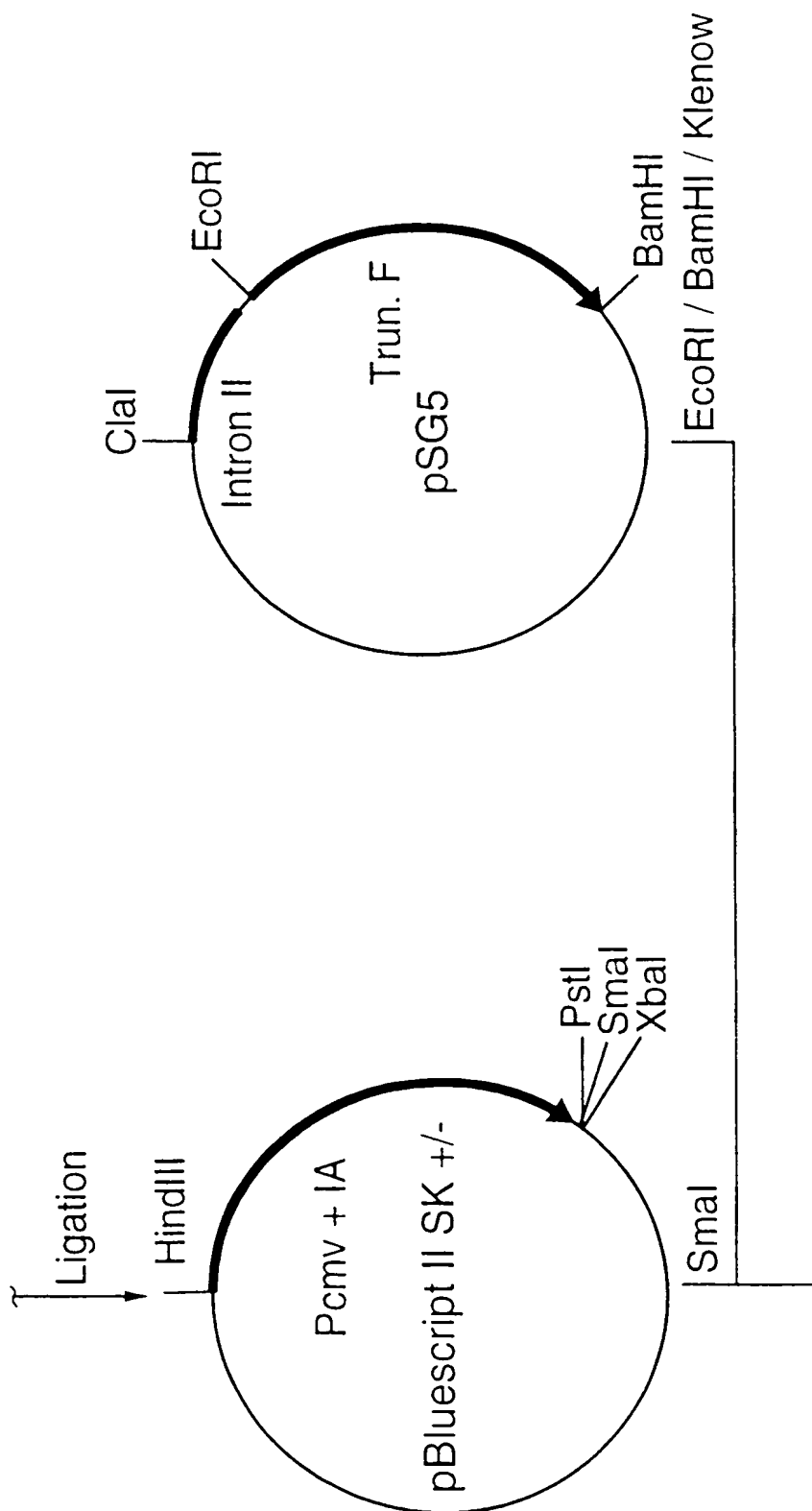


FIG.4B

07/9/4621

13/39

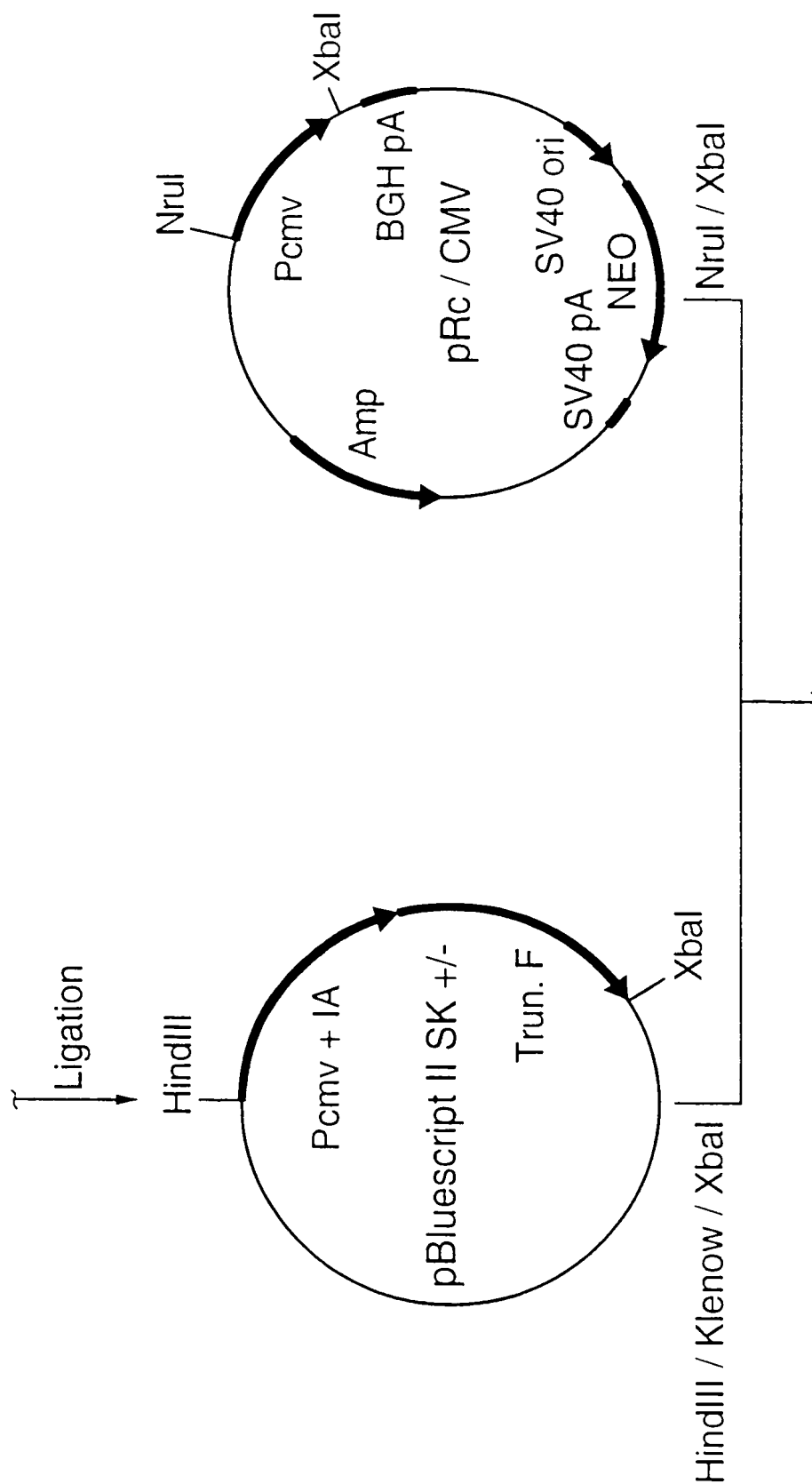


FIG.4C

09/9/4621

14 / 39

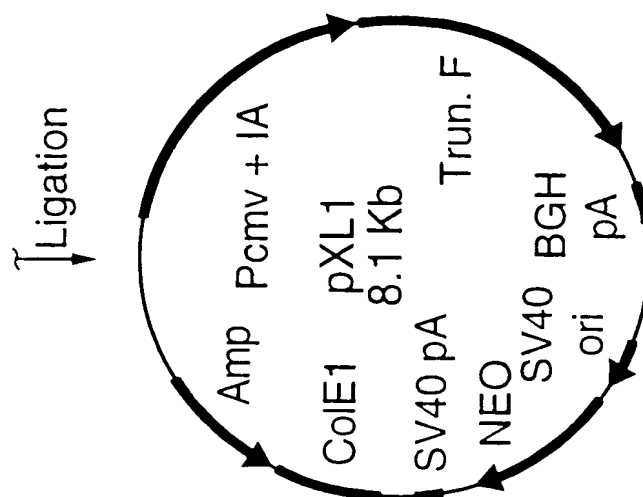


FIG.4D

15 / 39

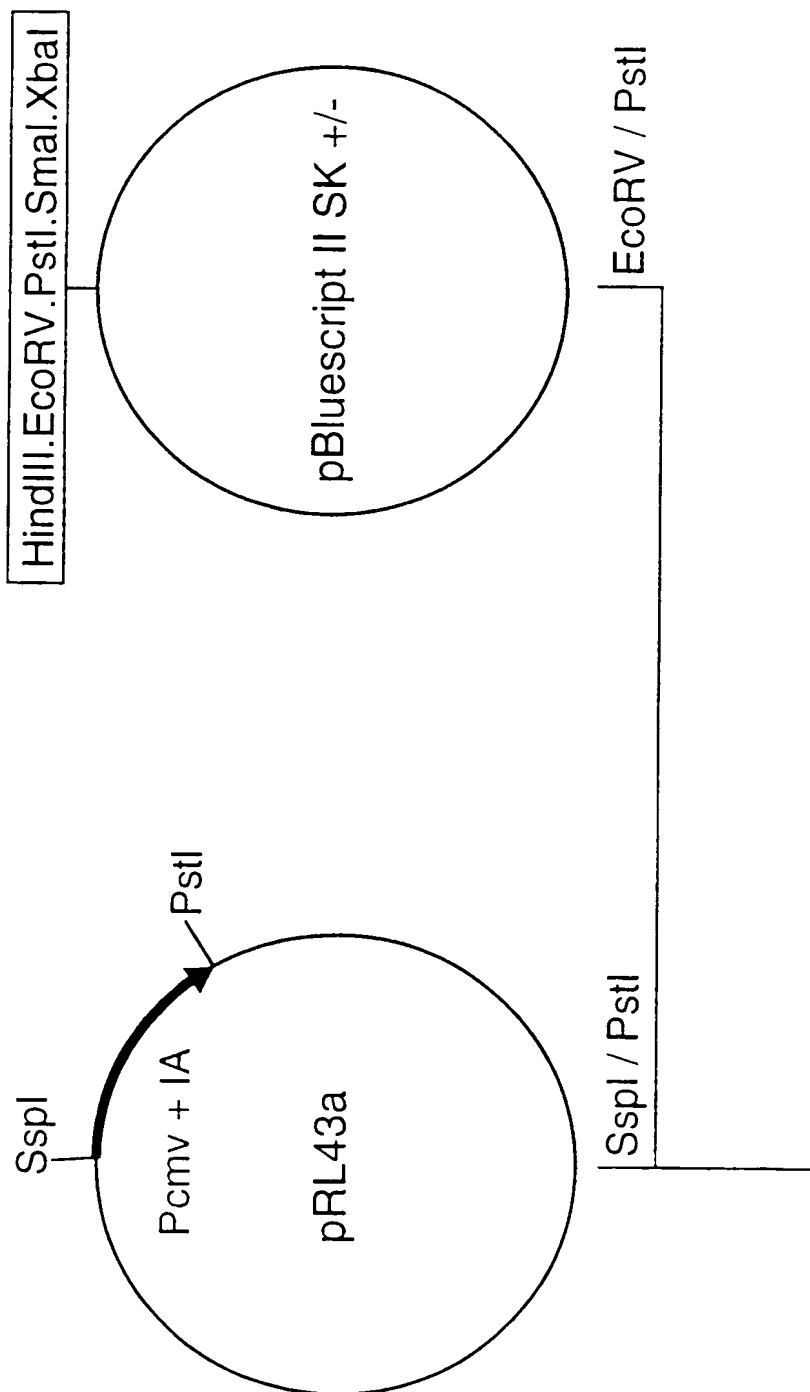


FIG.5A

16/39

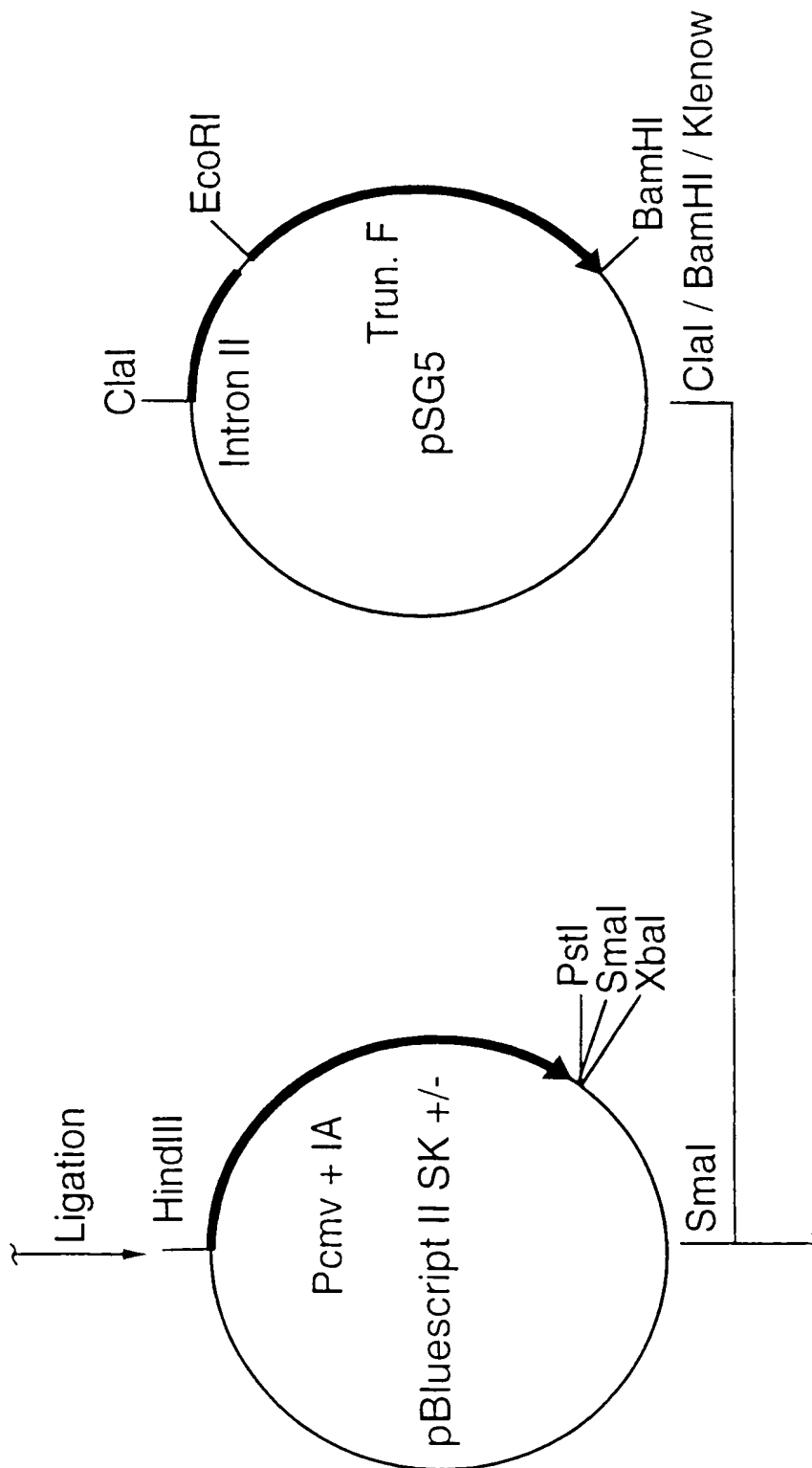


FIG.5B



17/39

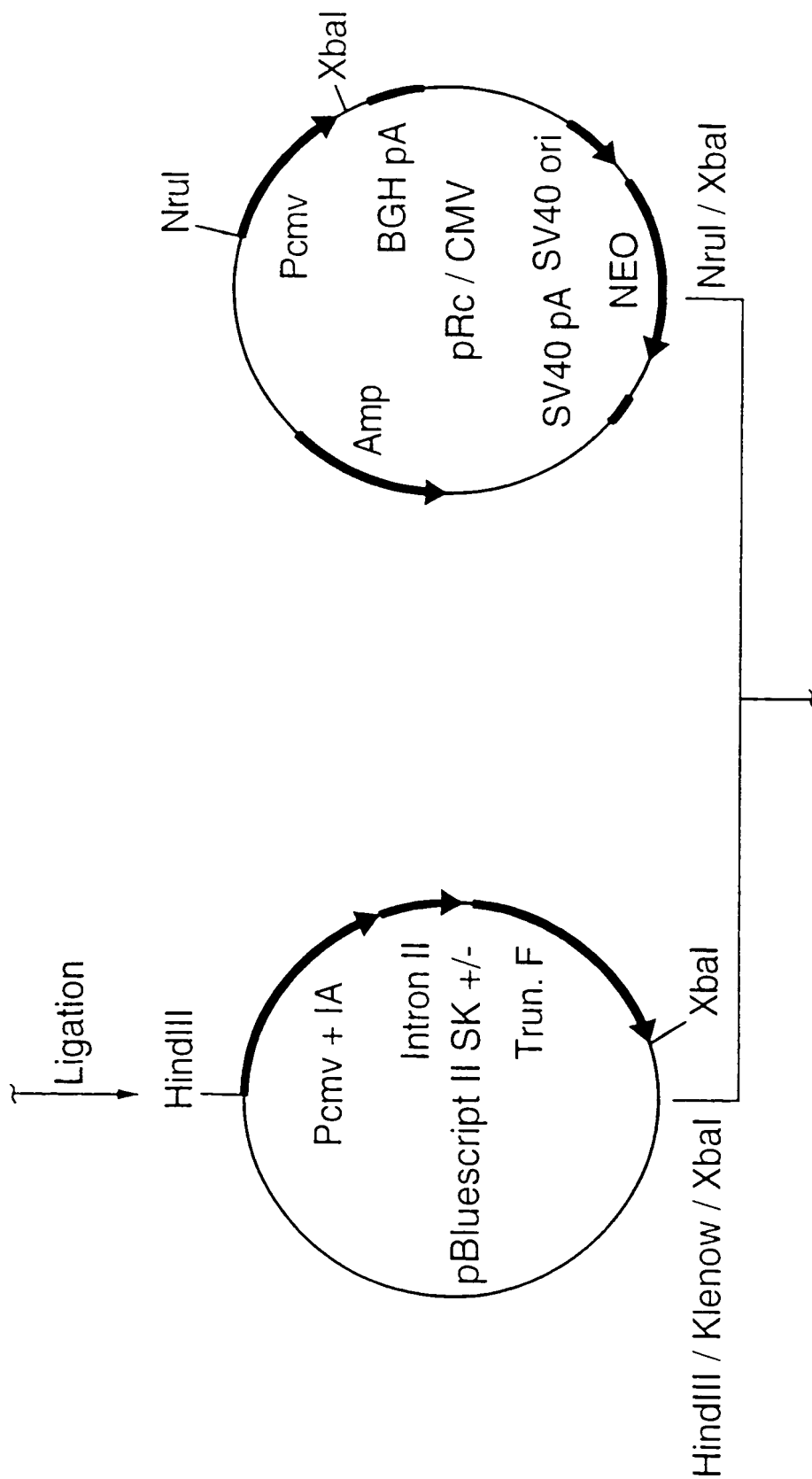


FIG.5C

09/9/4621

18/ 39

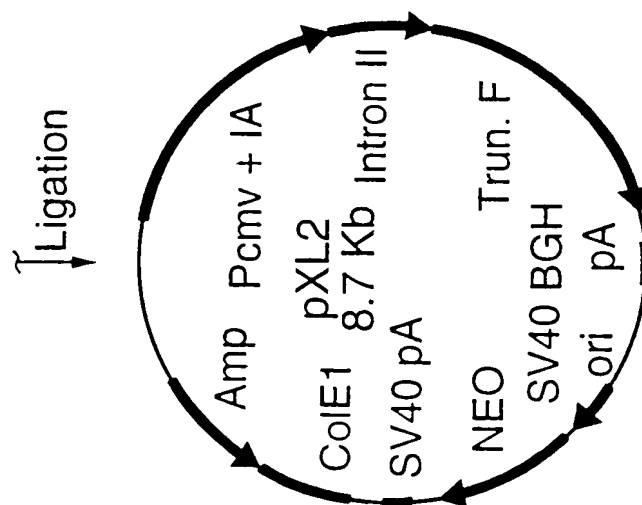


FIG.5D

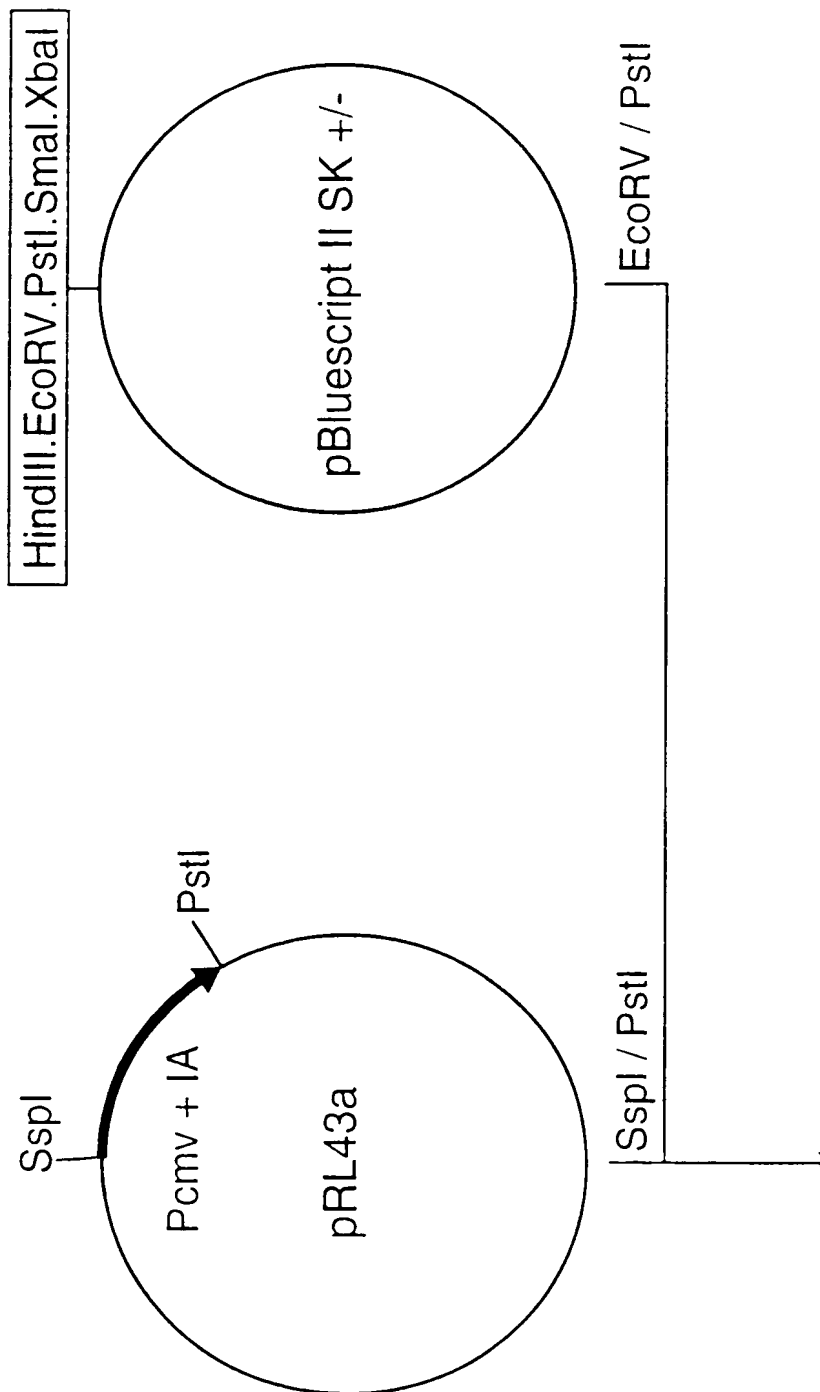


FIG.6A

09/91421

20/39

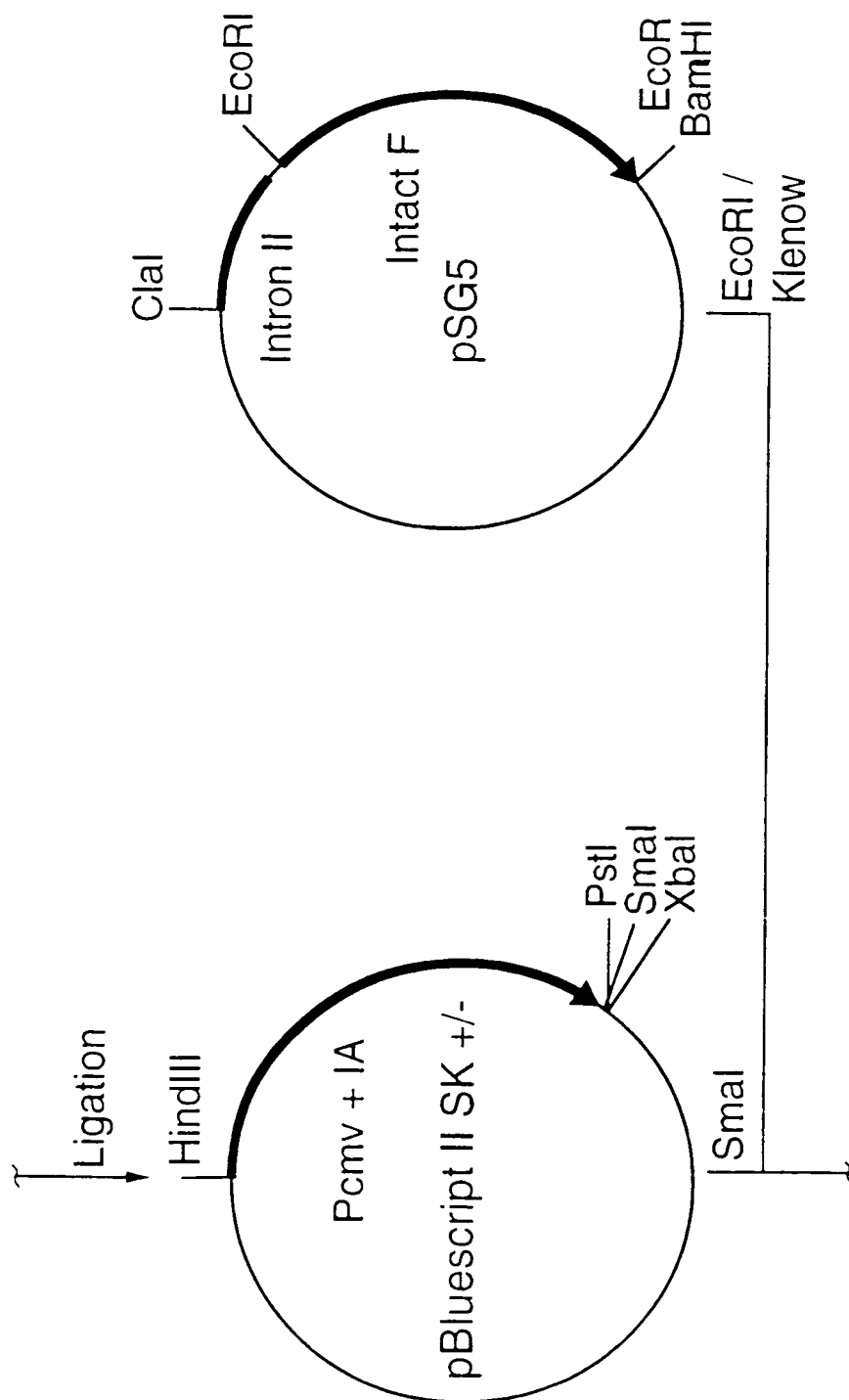


FIG.6B

09/9/4621

21 / 39

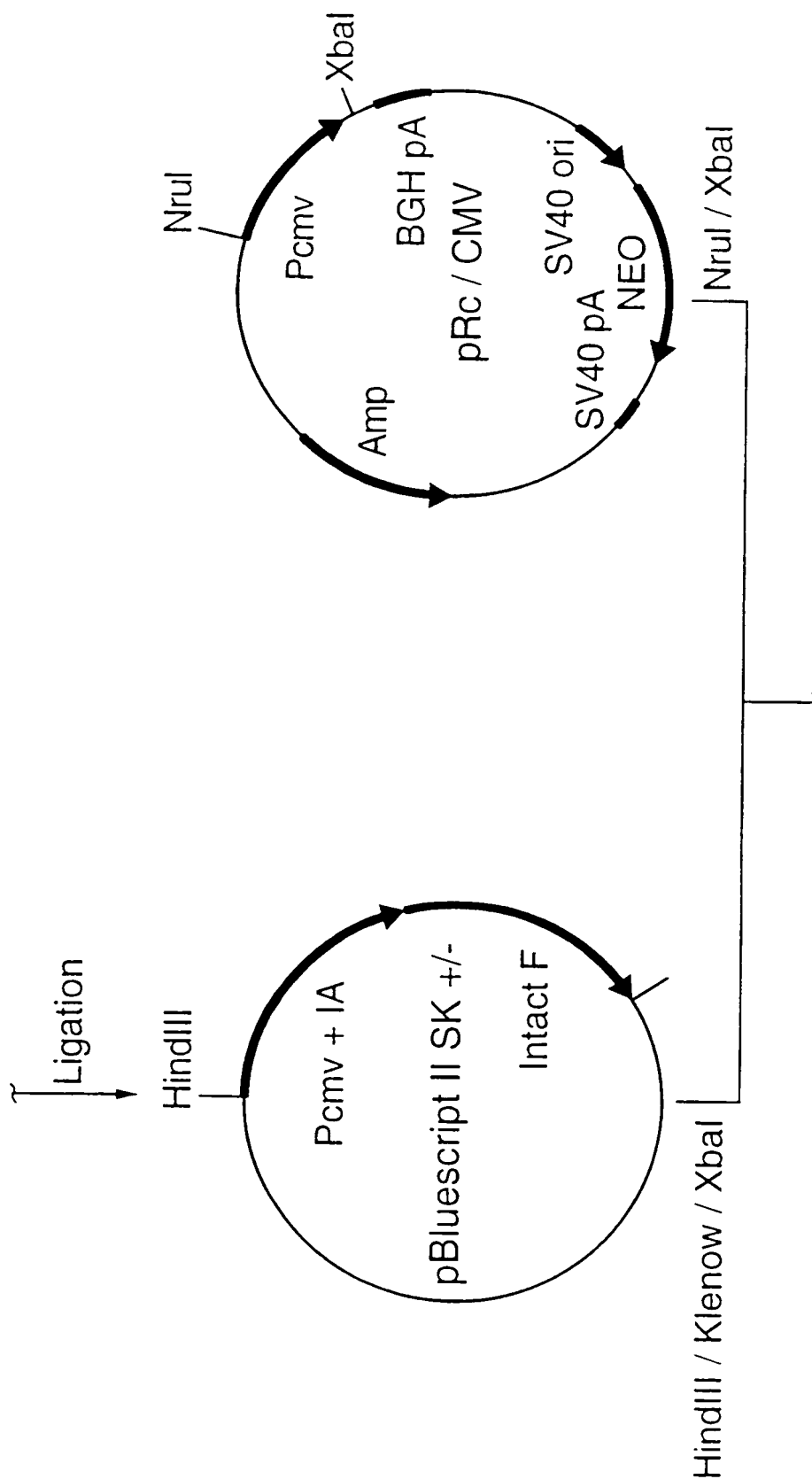


FIG.6C

22 / 39

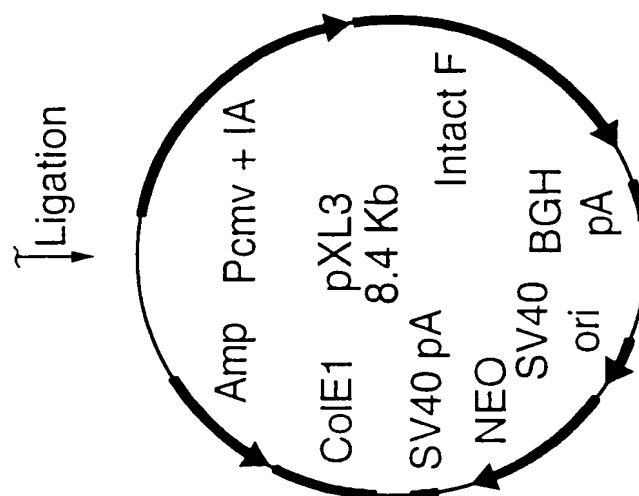


FIG. 6D

07/9/4621

23/39

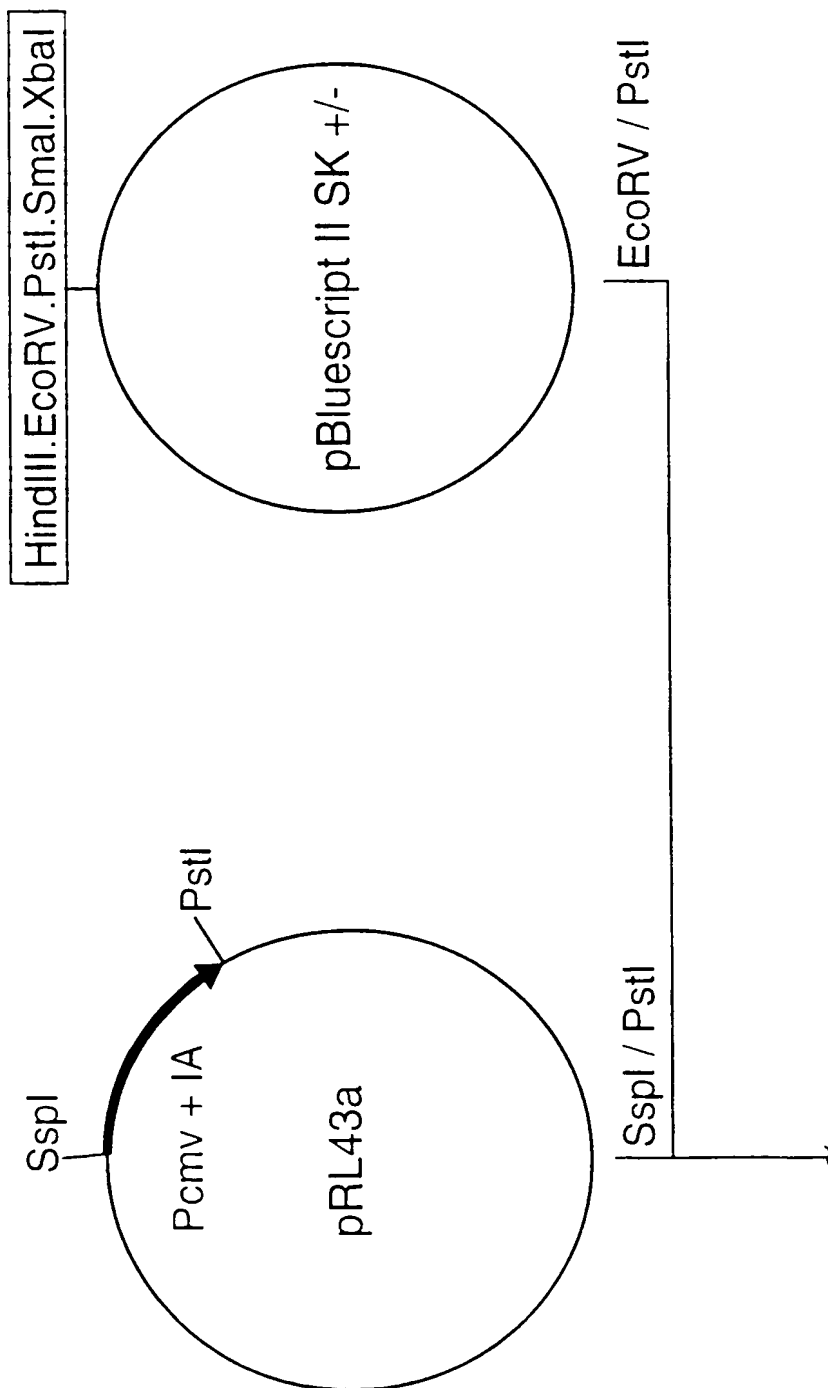


FIG.7A

09/9/4621

24/39

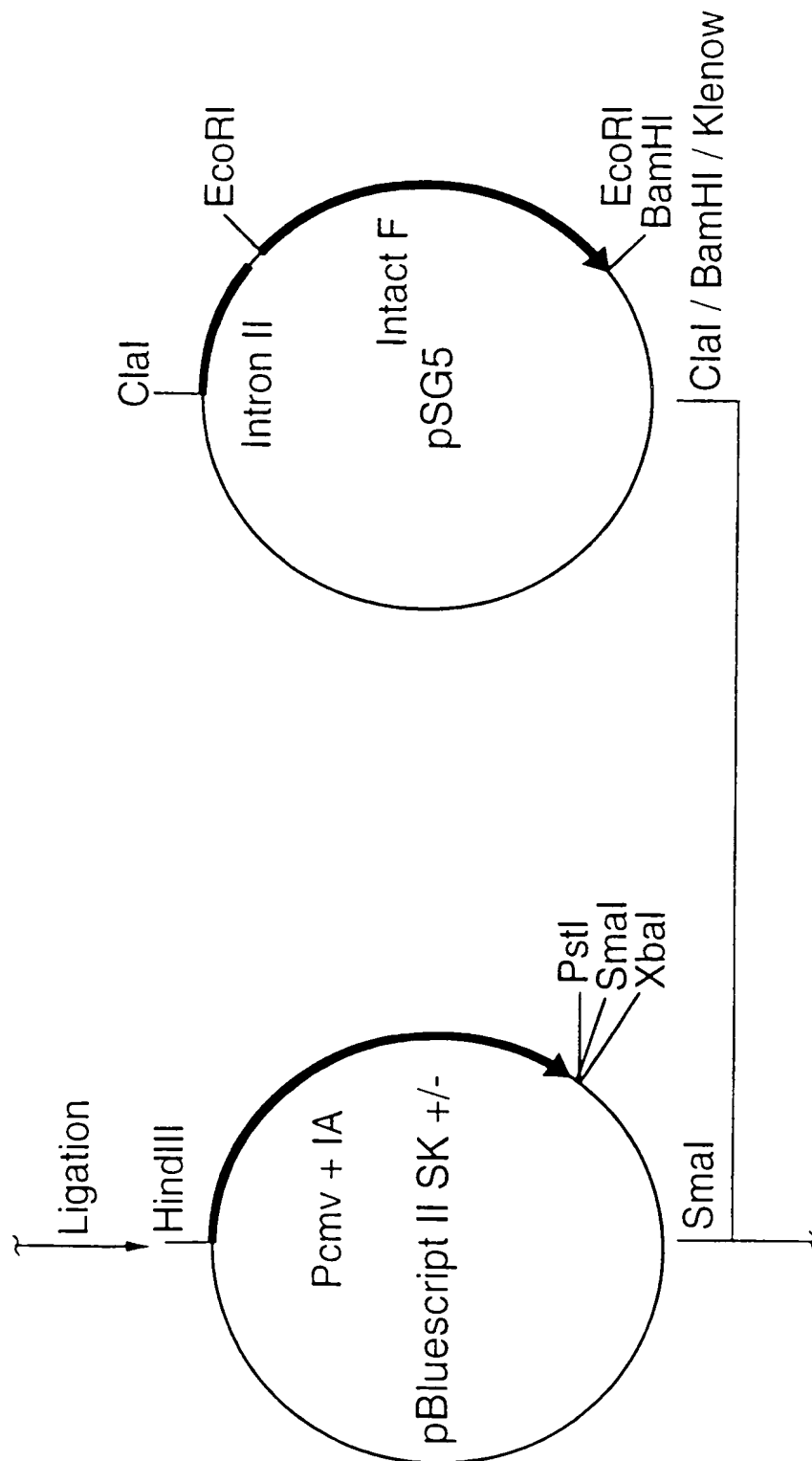


FIG.7B



09/9/4621

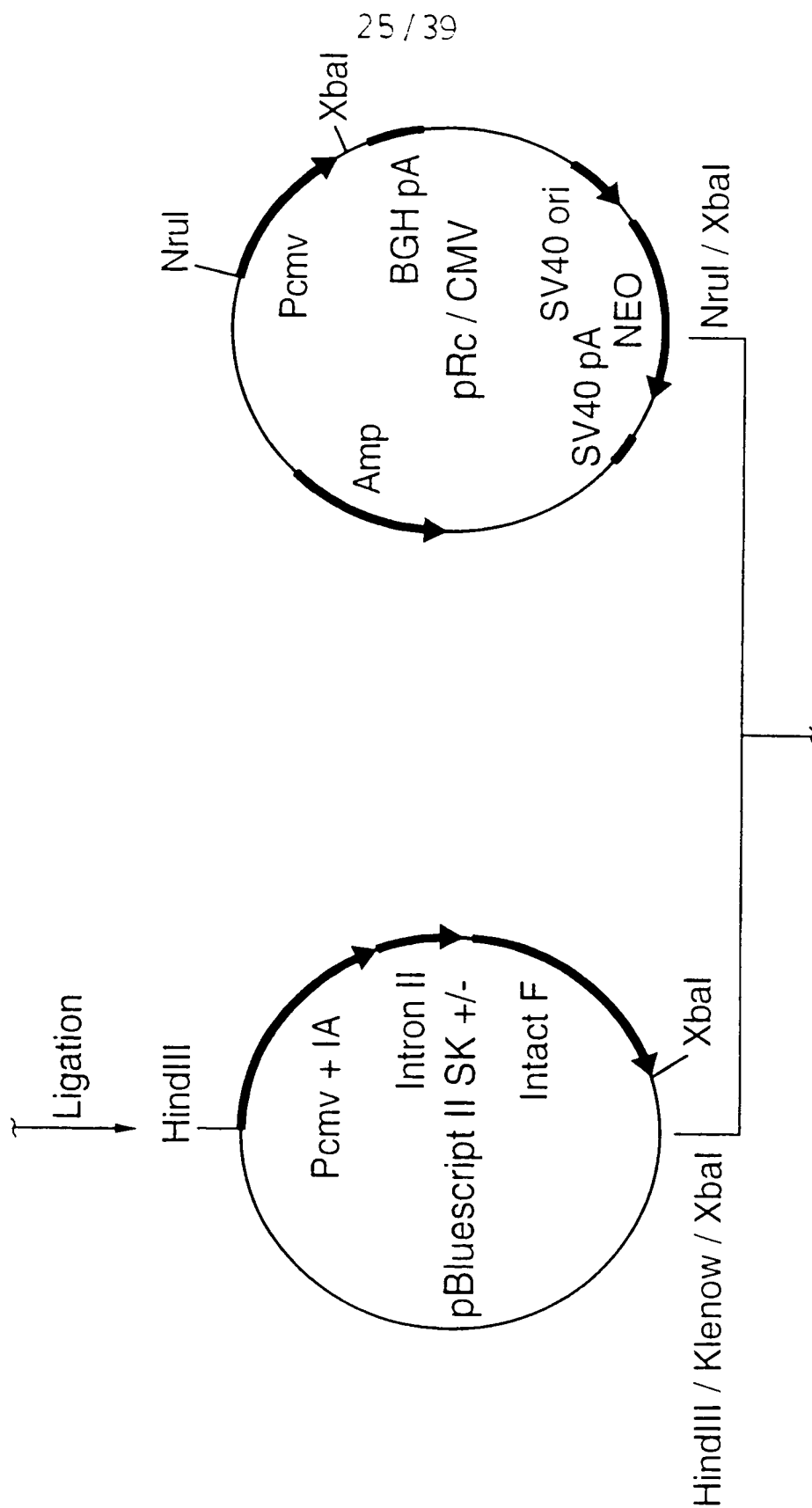


FIG.7C

26/39

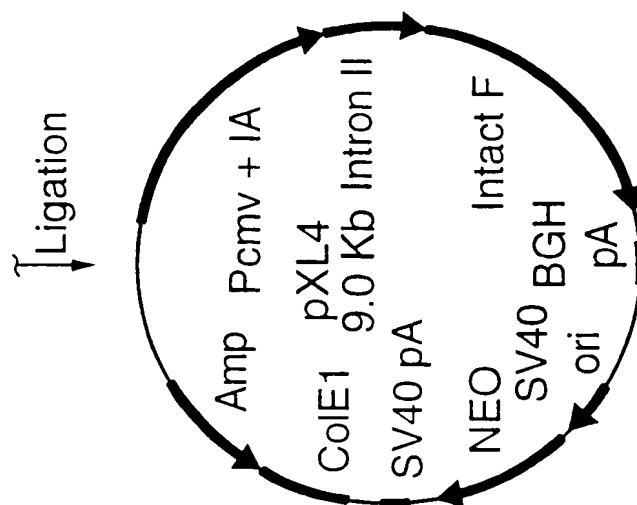


FIG.7D

27/39

FIG.8

GTGAGT  
401 TTGGGGACCC TTGATTGTTT TTTCTTTTTC GCTATGTAA AATTCATGTT  
451 ATATGGAGGG GGCAAAGTTT TCAGGGTGTTC GTTTAGAAATG GGAAGATGTC  
501 CCTTGTATCA CCATGGACCC TCATGATAAT TTTTGTTTCTT TCACATTCTTA  
551 CTCTGTTGAC AACCATGTC TCCCTCTTATT TTCTTTTTCAT TTTCTGTAAC  
601 TTTTTCGTTA AACTTTAGCT TGCATTGTA ACGAATTTT AAATTCACCT  
651 TTGTTTATTT GTCAGATTGT AAGTACTTTC TCTAATCACT TTTTTCCTCA  
701 GGCAATCAGG GTATATTATA TTGTAATTCA GCACAGTTT AGAGAACAAAT  
751 TGTATATAAT AATGATAAG GTAGAATATT TCTGCATATA AATCTGGCT  
801 GGCGTGGAAA TATTCTTATT GGTAAGAAACA ACTACATCCT GGTCAATCATC  
851 CTGCCCTTCT CTTTATGGTT ACAATGATAT ACACGTGTTG AGATGAGGAT  
901 AAAATACTCT GAGTCCAAAC CGGGCCCCCTC TGCTAACCAT GTTCATGCCCT  
951 TCTTCTTTT CCTACAG

28/39

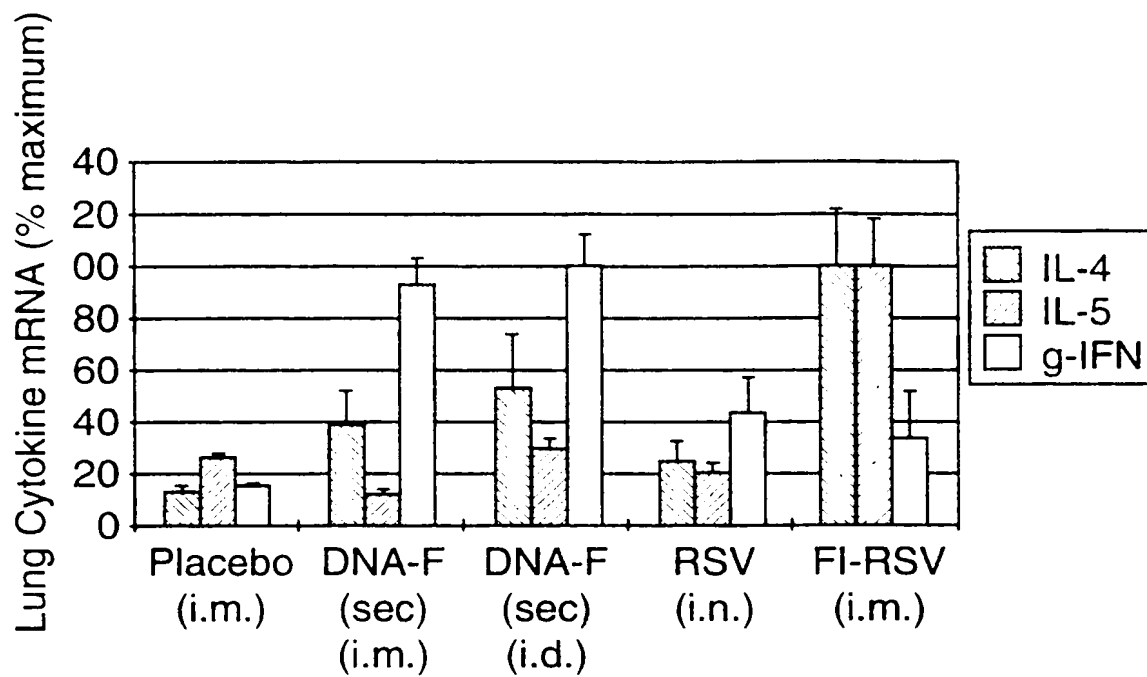


FIG.9

29/39

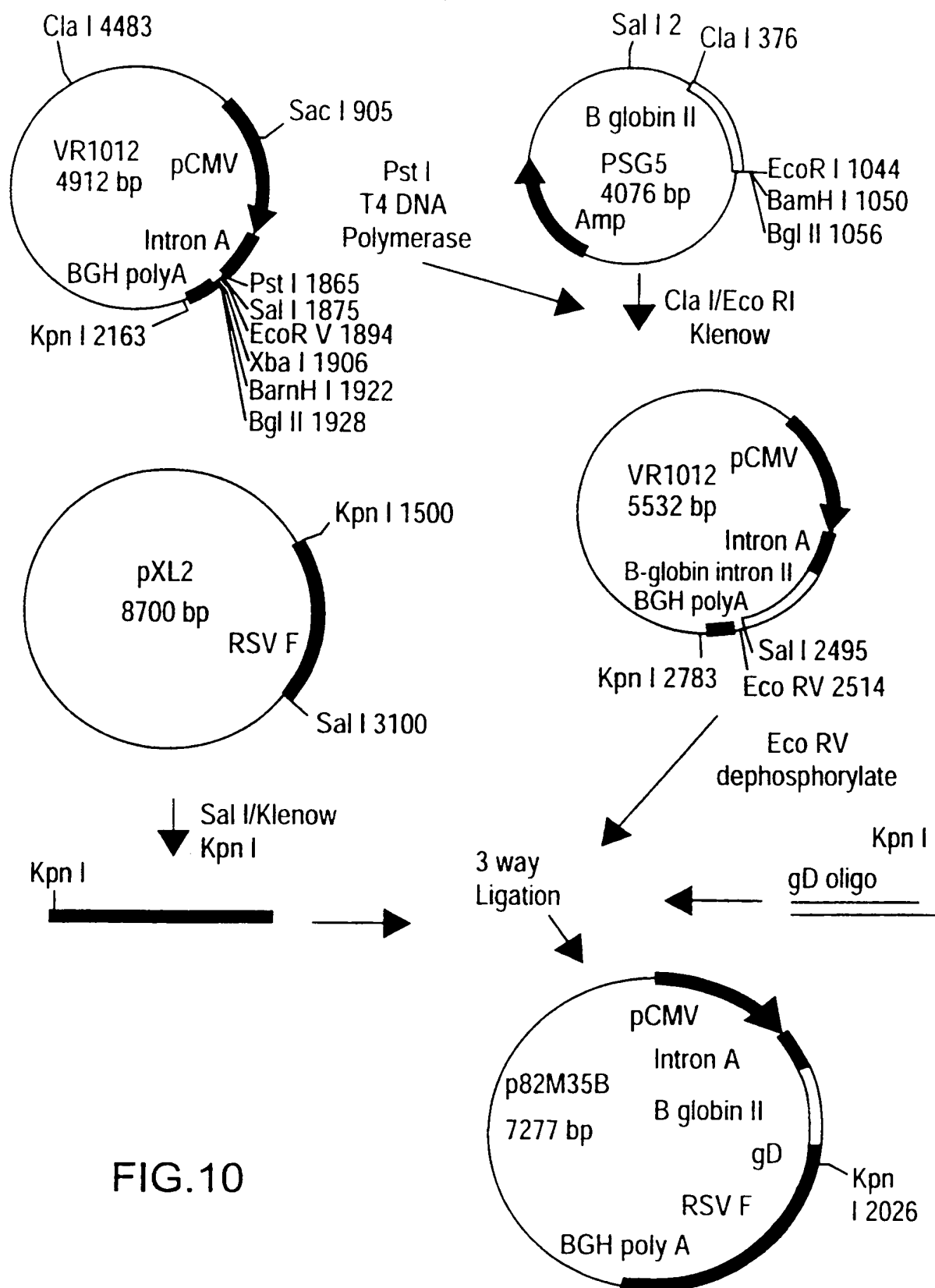


FIG.10

129416/160

## FIG.11A

## Nucleotide Sequence of plasmid VR1012

10 20 30 40 50 60 70  
TGGCGGGTTT OGGTATGAC GGIGAAAACC TCIGACACAT GCAGCTCCCG GAGACGGICA CAGCTTGTCT

80 90 100 110 120 130 140  
GTAAGCCGAT GCGGGAGCA GACAAGCCCG TCAGCGGCGG TCAGCGGGIG TTGGCGGGIG TCGCGGCTGG

150 160 170 180 190 200 210  
CTTAACIATG CCGCATCAGA GZAGATTGTA CTGACAGTGC ACCATATGCG GIGICAATA CCGCACAGAT

220 230 240 250 260 270 280  
GGGTACGAG AAAATACCGC ATCAGATTGG CTATTGGCCA TTGCATAAGT TGATATCATA TCATAATATG

290 300 310 320 330 340 350  
TACATTIATA TTGGCTCATG TCCAACATTA CCGCCATGTT GACATTGATT ATTGACIAGT TATTAAATAGT

360 370 380 390 400 410 420  
AATCAATTAC GCGGTCATTA GTTCAATAGCC CATATATCGA GTTCGGCGTT ACATAACTIA CCGTAAATGG

430 440 450 460 470 480 490  
CCCGCCCTGC TGACCGGCCC ACGACCCCGG CCGATTGACG TCAATAATGA CGTATGTTCC CATAGTAAAG

30/39

09/14/02

31/39

FIG.11B

500	510	520	530	540	550	560
CCAATAGCGA	CTTTCCATTTG	ACGTCATATGG	GTCGAGTATT	TACGGTAAAC	TGCCCATCTTG	GCATGTACATC
570	580	590	600	610	620	630
AAGGTATCA	TAAGCCAAAT	AGCCCCCTA	TTGACGTCAA	TGACGGTAAA	TGGCCCCCT	GGCATTTATGC
640	650	660	670	680	690	700
CCAGTACATG	ACCTTATGGG	ACTTTCCCTAC	TTGGCAGTAC	ATCTACGTAT	TAGTCATCGC	TATTACCAATG
710	720	730	740	750	760	770
GTCATCGGT	TTTGGCAGTA	CATCAATGG	CGTGGATAGC	GGTTTGACCTC	ACGGCGATTT	CCAAGTCTCC
780	790	800	810	820	830	840
ACCCCATTTGA	CGTCAATGGG	AGTTTGTTTT	GGCACCAAAA	TCAAACGGGAC	TTTCCAAAAT	GTGCTAACAA
850	860	870	880	890	900	910
CTCCGCCCCA	TTTACGGCAA	TGGCGGGTAG	GGGTGTACGG	TGGGAGGTCT	ATATAAGCAG	AGCTCGTTTA
920	930	940	950	960	970	980
GTCTAACGTC	AGATCGCCCTG	GAGACGCCAT	CCACGCTGTT	TTGACCTCCA	TAGAATACAC	CCGGTACCGAT
990	1000	1010	1020	1030	1040	1050
CCAGCCCTCG	CCGCGCGGAA	CCGTGCATTTG	GAACGGCGAT	TCCCGGTGCC	AAGAGTACAG	TAAGTACCGC

32/39

## FIG.11C

1060 1070 1080 1090 1100 1110 1120  
CTATAGACTC TATAGGCACA CCCCTTTGGC TCCTATGCGAT GCTATACIGT TTTTGGCTTG GGGCCATATAC

1130 1140 1150 1160 1170 1180 1190  
ACCCCCCGCTT CCTATAGCTA TAGGIGATGG TATAGCTTAG CCTATAGGIG TGGGTATIG ACCATTATIG

1200 1210 1220 1230 1240 1250 1260  
ACCACTCCCC TATTGGTGAC GATACTTTCC ATTACTAATC CATACATGG CTCCTTGGCA CAACTATCTC

1270 1280 1290 1300 1310 1320 1330  
TATTGGCTAT ATGCCAATAC TCCTGTCTTC AGAGACTGAC ACGGACTCTG TATTTTACG GGTATGGGTC

1340 1350 1360 1370 1380 1390 1400  
CCATTTTATTA TTATCAAAAT CACATATACA ACAAGCGCGT CCCCCGGTGC CGTAGTTTTT ATTAAACAATA

1410 1420 1430 1440 1450 1460 1470  
GGGTGGGATC TCCACGGGAA TCCTGGGTAC GGTTCGGG CAATGGGCTCT TCCTGGGTAG CCGGGGAGCT

1480 1490 1500 1510 1520 1530 1540  
TCCACATCG AGCCCTGGTC CCATGGCTCC AGCCGCTCAT GGTCGCTGG CAGCTGCTTG CTTCTAACAG

1550 1560 1570 1580 1590 1600 1610  
TGGAGGCCAG ACTTAGGATC AGCACAATCC CCACCACAC CAGTGTGG CACAAGGCCG TGGGGGTAGG



09/01/2021

33/39

## FIG.11D

1620	1630	1640	1650	1660	1670	1680
GTATGIGTCT	GAAATGAGC	GTCGAGATTG	GGCTGGCAGG	GCTGAGCCAG	ATGTAAGACT	TAAGGCAGCG
1690	1700	1710	1720	1730	1740	1750
GCAGAGAAG	ATGCAGGCAG	CTCAGTGTGT	GTATTCCTGAT	AAGAGTCAGA	GGTAACCTCC	GTTCGGGTCG
1760	1770	1780	1790	1800	1810	1820
TGTTAAACGGT	GGAGGGCAGT	GTAGTCTGAG	CAGTACTGGT	TGCTGGCGCG	CGGGCCACCA	GACATAATAG
1830	1840	1850	1860	1870	1880	1890
CTGCACAGACT	AACAGACTGT	TGCTTTCCAT	GGGTCTTTTC	TGCAGTCACC	GTCGTGCACA	CGTGTGATCA
1900	1910	1920	1930	1940	1950	1960
CAATATCGCG	CCGCTCTAGA	CCAGGGGCGT	GGATCCAGAT	CTGCTGTGCC	TTCCTAGTTGC	CAGCCATCTG
1970	1980	1990	2000	2010	2020	2030
TTGTTTGCC	CTCCCCCGTG	CCCTCCCTTGA	CCCTGGAAGG	TGCCACTGCC	ACGTGCTTTT	CCTAATAAAA
2040	2050	2060	2070	2080	2090	2100
TCAGCGAATT	GCATCGCAAT	GTCCTAGTAG	GTCCTATTCT	ATTCTGGGGG	GTCGGGTCGG	GCAGGACAGC
2110	2120	2130	2140	2150	2160	2170
AAGCGCGAGG	ATTTCGAAGA	CAATAGCAGG	CATGCTGGGG	ATCGGTCGGG	CTCTATGGGT	ACCCAGGTCG

09/9/4621

34/39

## FIG.11E

2180	2190	2200	2210	2220	2230	2240
TGAAGAATTG	ACCCGGTTCC	TCCCTGGGCA	GAAAGAAGCA	GGACATCCC	CTTCTCTGIG	ACACACCCIG
2250	2260	2270	2280	2290	2300	2310
TCCAGGCCCC	TGGTCTTAG	TTCAGCCCC	ACTCATAGCA	CACTCATAGC	TCAGGAGGCG	TCGGCCITCA
2320	2330	2340	2350	2360	2370	2380
ATCCACCCCG	CTAAAGTACT	TGGAGCGGTC	TCTCCCTCCC	TCATCAGCCC	ACCAAACCAA	ACCTAGCCTC
2390	2400	2410	2420	2430	2440	2450
CAACAGTGGG	AACAATTA	AGCAACATAG	GCTATTAACT	GCAGAGCGAG	AGAAATGCC	TCCAACATGT
2460	2470	2480	2490	2500	2510	2520
GAGGAAGTAA	TCAGAGAAAT	CATAGAAATT	CTTCGGCTTC	CTGGCTCACT	GACTGGCTGC	GCCTGGTGGT
2530	2540	2550	2560	2570	2580	2590
TOGGCTGGCG	CGAGCGGTAT	CAGCTCACTC	AAAGCGGGTA	ATACGGTAT	CCACAGATTC	AGGGCATTAAC
2600	2610	2620	2630	2640	2650	2660
GCAGCAACA	ACATGTGACC	AAAAGGCCAG	CAAAAGGCGA	GGTAACGTAA	AAAGCGCGCG	TTGCTGGCGT
2670	2680	2690	2700	2710	2720	2730
TTTTCCATAG	GCCTCGCCCC	CCGTAGGAGC	ATCACAAAA	TGCAGCTCA	AGTCAGAGGT	GGCGAAACCC

09/14/02

35 / 39

## FIG.11F

2740	2750	2760	2770	2780	2790	2800
GACAGGACTA	TAAAGATAAC	AGCGGTTC	CCCTGGAGC	TCCTCGTGC	GCCTCTCTGT	TCGACCCCTG
2810	2820	2830	2840	2850	2860	2870
CCGCTTAACG	GATACCTGTC	CGCCCTTCTC	CCCTCGGGA	GGGCGGGCT	TTCTCATAGC	TCACGCTGTA
2880	2890	2900	2910	2920	2930	2940
GGTATCTCAG	TTCGGGTAG	GTCGTTCCT	CCAGGCTGG	CTGGTGGAC	GAACCCCGG	TTTAGCCCCGA
2950	2960	2970	2980	2990	3000	3010
CCGCTGGGOC	TTATCCGGTA	ACATATGCT	TCAGTCCAAC	CCGGTAAGAC	ACGACTTATC	GCCTACTGGCA
3020	3030	3040	3050	3060	3070	3080
GCAGCCACTG	GTAAACAGGAT	TACGACAGCG	AGGTATGTAG	GGGGTCTAC	AGAGTCTTTC	AAGTGGTGGC
3090	3100	3110	3120	3130	3140	3150
CTAACTAOCG	CTACACTAGA	AGTAACAGTAT	TGGTATCTG	CGCTCTGCTG	AAGCCAGTTA	CCCTCGGAAA
3160	3170	3180	3190	3200	3210	3220
AAGAGTTCGT	AGCTCTTGAT	CCGGCAACA	AACACCGCT	GGTAGGGTTC	GTCTTTTCTG	TTGCAAGCTAG
3230	3240	3250	3260	3270	3280	3290
CAGATTACGC	GCACAAAAA	AGGATCTCAA	GAAGATCCTT	TCATCTTTTC	TACGGGGTCT	GACGCTCAGT

09/14/02

36/39

## FIG.11G

3300 3310 3320 3330 3340 3350 3360  
GGAACGAAA CTCACGTTAA GGGATTTCG TCATGAGATT ATCAAAAAGG ATCTTCACCT AGATCCCTTT

3370 3380 3390 3400 3410 3420 3430  
AAATTAAAA TGAAGTTTAA AATCAATCTA AAGTATATAT GAGTAAACTT GGTCIGACAG TTACCAATGC

3440 3450 3460 3470 3480 3490 3500  
TTAATCAGTG AGCCACCTAT CTCAGCGATC TGCTATTTC GTTCATCCAT AGTTCCTTGA CTGGGGGGGG

3510 3520 3530 3540 3550 3560 3570  
GGGGGGGCTG AGGTCIGCCCT CGTCAAGAAG GGTTCGCTGA CTCATACCAG GCGTGAATCG CCGCATCATC

3580 3590 3600 3610 3620 3630 3640  
CAGCCAGAAA GTCAGCGAGC CACGGTTGAT GAGAGCTTTC TGTGAGGTGG ACCAGTTGGT GATTTTGAC

3650 3660 3670 3680 3690 3700 3710  
TTTTTCCTTG CCACCGAAGG GTCIGCGTTG TCGCGAAGAT GCGTGTATCTG ATCTTCAAC TCAGCAAAAG

3720 3730 3740 3750 3760 3770 3780  
TTTCGATTAT TCACCAAAGC GCGCGTCCCG TCATAGTCAGC GTATGCTCT GCGAGTGTGA CAACCAATTA

3790 3800 3810 3820 3830 3840 3850  
ACCAATTGTG ATTAGAAAAA CTCATCGAGC ATCAAAATGAA ACTGCAATT ATTCAATATCA GCAATTATCAA

37/39

FIG.11H

3860            3870            3880            3890            3900            3910            3920  
TACCATATTT TTGAAAAGC CGTTTCTGTA ATGAAGGAGA AAATCACCAG AGGAGATTCC ATAGGATGGC

3930            3940            3950            3960            3970            3980            3990  
AAGATCCCTGG TATCGGTCIG CGATTCCGAC TCGTCCCAACA TCAATACAAC CTATTAAATTT CCCCTCGTCA

4000            4010            4020            4030            4040            4050            4060  
AAAATAAGGT TATCAAGTGA GAAATCACC A TGAGTGAAGA CTGAATCCGG TCAGAAATGGC AAAAGCTTAT

4070            4080            4090            4100            4110            4120            4130  
GCATTTCITTT CCAGACTTGT TCAACAGGCC AGCATTACG CTGCTCATCA AAATCACCTGG CATCAACCAA

4140            4150            4160            4170            4180            4190            4200  
ACCGTTATTC ATTGGTGATT GCGCCCTGAGC GACAGCAAAAT ACGCGATCGC TGTTAAAGG ACAATTACAA

4210            4220            4230            4240            4250            4260            4270  
ACAGGAATCG AATGCAACCG GGCAGGAAC ACIGCCAGCG CATCAACAAT ATTTTCACTT GAATCAGGAT

4280            4290            4300            4310            4320            4330            4340  
ATTCTCTTAA TACCTCGAAT GCCTGTTTCC CCGGATCGC AGTGGTGAAT AACCATGCAT CATCAGGAGT

4350            4360            4370            4380            4390            4400            4410  
ACGGATAAAA TGCTTGATGG TCGGAAGAGG CATAAATTCC GTCAGCCAGT TTAGTCTGAC CATCTCATCT

09/11/90

## FIG. 11I

4420	4430	4440	4450	4460	4470	4480
GTAACATCAT	TGGCAACGCT	ACCCTTGCCA	TGTTTCAGAA	ACAACCTCTGG	CGCATCGGCG	TTCOCATACA
4490	4500	4510	4520	4530	4540	4550
ATCGATAGAT	TGTGCGACCT	GATTGCGCGA	CATTATCGCG	AGCCCATTTA	TACCCATATA	AATCAGCATC
4560	4570	4580	4590	4600	4610	4620
CATGTTGGAA	TTTAATCGCG	GCCTCGAGCA	ACAAGTTTCC	CGTTCAATAT	GGCICATAAC	GTTCCTTGTA
4630	4640	4650	4660	4670	4680	4690
TTACTGTGTTA	TGTAAGCACA	CAGTTTATTT	GTTCATCATG	ATATATTTTT	ATCTTGTCGA	ATGTAACATC
4700	4710	4720	4730	4740	4750	4760
ACAGATTITG	AGACACAACG	TGGCTTTCCC	CCCCCCCCCA	TTATIGAAGC	ATTATCAGG	GTATTTGICT
4770	4780	4790	4800	4810	4820	4830
CATGAGCGGA	TACATATTTG	AATGTATTTA	GAAAAATAAA	CAAAATAGGG	TTCGGGCGAC	ATTTCOCCTA
4840	4850	4860	4870	4880	4890	4900
AAAGTGCCAC	CTGACGTCTA	AGAAACCATT	ATTATCATCA	CATTAACTTA	TAAAAATAGG	CGTATCAGCA
4910	GGCCCTTTGG TC					

38/39

09/9/4621

FIG. 12

5'AAG CTT CAG GAA CGA CCA ACT ACC CCG ATC ATC AGT TAT CCT  
TAA GGT CTC TTT TGT GTG GTG CGT TCC GGT ATG GGG GGG ACT GCC  
GCC AGG TTG GGG GCC GTG ATT TTG TTT GTC GTC ATA GTG GGC CTC  
Ala Arg Leu Gly Ala Val Ile Leu Phe Val Val Ile Val Gly Leu  
CAT GGG GTC CGC GGC AAA TAT GCC TTG GCG GAT GCC TCT CTC 3'  
His Gly Val Arg Gly Lys Tyr Ala Leu Ala Asp Ala Ser Leu

39/39

Met Gly Gly Thr Ala